R E P O R T R E S U M E S

ED 015 081 RE 000 216

A TWO-YEAR LONGITUDINAL STUDY TO DETERMINE THE ABILITY OF FIRST GRADE CHILDREN TO LEARN TO READ USING THE EARLY-TO-READ I/T/A PROGRAM.

BY- MCCRACKEN, ROBERT A.
WASHINGTON OFF. STATE SUPT. PUB. INSTR., OLYMPIA

REPORT NUMBER RR-07-07-1967 PUB DATE JUN 67
EDRS PRICE MF-$0.50 HC-$3.08 75P.

DESCRIPTORS- *INITIAL TEACHING ALPHABET, *READING INSTRUCTION, *BEGINNING READING, *GRADE 1, *GRADE 2, READING TESTS, BASIC READING, WESTERN WASHINGTON STATE COLLEGE, BELLINGHAM, WASHINGTON, MUKILTEO SCHOOL DISTRICT 36, MUKILTEO, OFFICE OF THE SUPERINTENDENT OF PUBLIC INSTRUCTION, OLYMPIA,

A 2-YEAR STUDY OF THE ABILITY OF FIRST GRADERS TO LEARN TO READ USING THE "EARLY-TO-READ I/T/A/" SERIES WAS CONDUCTED. SIXTY-ONE FIRST GRADERS IN ROSE HILL ELEMENTARY SCHOOL, MUKILTEO, WASHINGTON, WERE RANDOMLY ASSIGNED TO AN EXPERIMENTAL GROUP USING THE I/T/A SERIES AND TO A CONTROL GROUP USING THE "GINN BASIC READERS." THREE RANDOMLY CHOSEN FIRST-GRADE CLASSES SERVED AS THE SUBCONTROL GROUP. ACCORDING TO READINESS TESTS, AN INTELLIGENCE SCALE, AND AN ALPHABET KNOWLEDGE TEST, THE EXPERIMENTAL AND CONTROL GROUPS WERE NOT SIGNIFICANTLY DIFFERENT. ABOUT 160 BEGINNING READING TRADE BOOKS WERE TRANSLITERATED FOR THE EXPERIMENTAL GROUP. TRADITIONAL EDITIONS WERE MADE AVAILABLE TO THE CONTROL GROUP. OTHER TESTS GIVEN WERE THE GRAY ORAL READING TESTS, THE STANDARD READING INVENTORIES, AND THE STANFORD ACHIEVEMENT TESTS. RESULTS INDICATED THAT THERE WERE NO SIGNIFICANT DIFFERENCES IN FIRST-GRADE ACHIEVEMENT, IN SECOND-GRADE ACHIEVEMENT, OR IN OVERALL READING ACHIEVEMENT BETWEEN THE EXPERIMENTAL AND CONTROL GROUPS. BOTH GROUPS WERE SUPERIOR TO THE SUBCONTROL GROUP. THE EXPERIMENTAL GROUP READ SIGNIFICANTLY BETTER THAN THE OTHER TWO GROUPS WHEN READING I/T/A MATERIALS AND CONSISTENTLY SCORED HIGHEST. THE SUBCONTROL GROUP CONSISTENTLY SCORED LOWEST. (NS)

RE 000 216

A TWO-YEAR LONGITUDINAL STUDY TO DETERMINE THE ABILITY OF FIRST GRADE CHILDREN TO LEARN TO READ USING THE EARLY-TO-READ i/t/a PROGRAM

ED015081

Robert A. McCracken
Reading Center Director
Western Washington State College
Bellingham, Washington 98225

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RE 000 216

This research was conducted jointly by Mukilteo School District #6, Mukilteo, Washington, Mr. J. O. Simpson, Superintendent of Schools, and Western Washington State College. This research was funded through the Research Division of the Office of the Superintendent of Public Instruction of the State of Washington.

LOUIS BRUNO
State Superintendent of Public Instruction
Olympia, Washington

Research Report 07-07
June 1967

TITLE: A TWO-YEAR LONGITUDINAL STUDY TO DETERMINE THE ABILITY OF FIRST GRADE CHILDREN TO LEARN TO READ USING THE EARLY-TO-READ i/t/a PROGRAM

In 1960 John A. Downing, Reading Research Officer at the University of London, initiated experimental work using the New Augmented Roman Alphabet[1] in teaching beginning reading. This is commonly called i/t/a, Initial Teaching Alphabet. I/t/a has 43 letters. These letters are identical with traditional orthography insofar as possible, and the added letters retain a similarity to traditional orthography. Spelling in i/t/a is regularly phonetic, with a few exceptions so that the transition from i/t/a to traditional orthography can be made easily. For example, the ck digraph representing the phoneme k has been retained. However, c is never used to represent the phoneme s.

Children seem to learn to read very quickly in i/t/a, and have been reported as changing to traditional orthography at approximately mid-first grade without instruction or difficulty.[2] Later reports suggest that transfer should not be pushed but might take two or more years.

Mazurkiewicz and Tanyzer[3] in 1963 published a completely new set of readers and workbooks for teaching reading in the United States by use of i/t/a. The series includes seven readers, six workbooks, teachers' manuals and supplementary teaching aids. The American version of i/t/a has 44 letters.

Mazurkiewicz is directing a six-year longitudinal study using his own materials, The Early-to-Read i/t/a Program, in Bethlehem, Pennsylvania.

Mazurkiewicz[4] reported on the Bethlehem experiment at the end of grade two. Comparisons were made using the California Reading Test, Upper Primary Form and the Stanford Achievement Tests, Primary II. (Mazurkiewicz does not

report the form of either test.) When comparing the total population Mazurkiewicz reported no differences in vocabulary or comprehension on the California Reading Test and no differences on the Stanford Achievement Test except in spelling. He reports a mean for the i/t/a group of 16.9 and a mean for the t.o. group of 13.8 on the spelling test. The t-ratio of 2.56 is noted as "significant at close to the 1 percent level." Mazurkiewicz notes that the i/t/a taught children are reading from higher reader levels in their classes and in comparing writing samples reports that the i/t/a taught children are generally superior ($P < 0.01$).

At the end of three years Mazurkiewicz[5] reports:

"The three year evaluation shows that children in i/t/a materials

1. learn more readily, achieve significantly superior reading skill at an earlier time, read more widely, write more prolifically, extensively, and with a higher degree of proficiency than their t.o. counterparts and have no difficulty in making a reading transition to t.o. materials when they are allowed to develop the appropriate confidence and efficiency levels.
2. spelling skill in i/t/a (better described as encoding) approaches perfection fairly early. The transition to spelling in t.o. is relatively easy when directed instruction and guidance is given in subsequent years and achievement in spelling on standardized tests and in creative writing significantly better than the achievement of t.o.-taught children at the end of the second and third years.
3. word recognition achievement in t.o. at the end of the first and second years is significantly better for i/t/a-taught children but no differences are found at the end of the third year from the t.o.-taught population.

4. the lack of inhibition in writing for i/t/a-taught children noted early in the first year continues unabated into the second and third years and significant differences are found in these children's creative writing in terms of the number of running words and the number of polysyllabic words used. No differences in the use of the mechanics of reading were found between the populations.
5. comprehension as measured by instructional levels and reader level achievement in all years favor the i/t/a population, while standardized test achievement in comprehension shows no differences from the t.o. population.
6. no inferior results on such measures as rate of reading or accuracy of reading are found, suggesting that the i/t/a to t.o. procedure establishes no negative characteristics, no hindrances on later achievement."

The i/t/a Bulletin[6] has reported success in each issue for school systems using i/t/a for initial reading instruction and remedial instruction.

It seemed desirable to check the value of the Early-to-Read i/t/a Program, independent of its author, since it was the only i/t/a program available in the United States in 1964 although three or more programs were available in England in September 1964.

The U.S. Office of Education first grade reading studies [7,8,9,10,11] included five studies comparing i/t/a with a variety of other approaches to beginning reading. No study reported i/t/a as yielding achievement results which were significantly different from all of the other methods tested in each study. Hayes and Tanyzer each reported i/t/a as yielding significantly better

results than traditional basic reading but not significantly different than the Lippincott series which was used as one of the comparative methods. Hayes did note that i/t/a seemed to be better for students with low IQ's.

Jones[12] reported a second British experiment replicating Downing's previous work. Jones found that the i/t/a taught children were generally superior to those taught with t.o. when the testing was done in i/t/a and t.o. respectively. She found no significant differences when both groups were tested in t.o. On the Neale Analysis of Reading Ability, Form C Jones reports means for the experimental children of 16.69 for accuracy, 21.95 for rate, and 5.35 for comprehension when tested in i/t/a. For the control children tested in t.o. she reports means for accuracy of 11.30, for rate 19.49, and comprehension of 4.39. On the Neale Analysis of Reading Ability, Form A administered in t.o. to both groups, Jones reports the following means: for the experimental group, accuracy 12.39, rate 21.17, comprehension 4.65; for the control group, accuracy 13.20, rate 22.44, comprehension 4.69. Jones reports the mean achievements of 16 schools. When the means of the Neale Test administered in i/t/a and t.o. are compared, all 16 experimental groups lost or regressed. On the same comparison 15 of the 16 control schools increased in score. If we used a sign test to compare the significance of these shifts, both would be significant ($P<0.01$). The shift for the control group could reflect differences between the two forms of the test or it could be that the second form score was better because something was learned in the first testing situation. Since the testing experiences were the same, we would expect that the same favorable factors would be working for the experimental group. Therefore, the regression may be even greater than appears.

Dunn, Mueller and Neely[13] investigated the efficacy of the Peabody

Language Development Kit and the initial teaching alphabet in teaching beginning reading. Children worked in four groups, those who had i/t/a plus PLDK, those with i/t/a only, and those who had PLDK only, and the control group. All three experimental groups scored higher on Metropolitan School Achievement Tests than the control group. The i/t/a plus PLDK group scored highest, i/t/a only next, and PLDK only third. Differences were significant between all groups except there was no significant difference between the two i/t/a groups.

Purpose:

The primary purpose of this study was to evaluate the Early-to-Read i/t/a Program. The secondary purposes were to evaluate the individual's rate of learning to read and to examine the relationship of intelligence to success in learning to read under the Early-to-Read i/t/a Program.

Initiation of the i/t/a Teaching Experiment:

The i/t/a project was initiated by Mr. Lawrence Ames, Principal in the Mukilteo School District #6. Mr. Ames had been assigned to the Rose-Hill Elementary School for the 1964-65 school year. It seems important to note that this was not an experiment initiated by teachers. The Rose-Hill Elementary School houses two first grades. Both first grade teachers expressed a willingness to cooperate in the experiment. The experiment was designed and a summer meeting was held on August 23, 1964 to explain the program to parents.

The subjects:

Sixty-one first grade pupils entered the Rose-Hill Elementary School in September of 1964. Thirty-four of these pupils were randomly assigned to the experimental i/t/a group and twenty-six of the pupils were randomly assigned to the traditionally taught control group. The larger number of

children was assigned to the i/t/a group because it was felt that pupils moving into the Rose-Hill Elementary School District during the year would need to be placed in the traditionally taught classroom and that by the end of the school year the classes would be approximately the same size. It was felt that drop-outs might cause a loss of five to ten pupils in either class and that the drop-out loss would be a more severe problem in the i/t/a class. It was felt that the i/t/a experimental class should be large enough so that the results achieved would not be attributed to a small class size. The fear of subject loss proved unwarranted. The i/t/a class lost only one pupil during the first school year and three during the second. The control class lost one pupil before any testing was done and seven during the second year.

Three first grade classes were chosen randomly from the remaining 10 first grades in the Mukilteo School District #6 to serve as a second control group. This group is referred to in this report as the sub-control group. These classes received only initial readiness testing and final reading achievement testing at the end of grades 1 and 2. It was felt that a Hawthorne effect might be present in both the i/t/a and control groups in the Rose-Hill School since the children would be tested individually throughout the year and the classes would be observed regularly throughout the year.

The two Rose-Hill teachers were randomly assigned to the i/t/a experimental class and the t.o. control class. The children in the i/t/a experimental class and the t.o. control class remained as separate classes during second grade. It was planned to randomly assign half of the i/t/a and half of the t.o. children to each of the two second grades. It was not possible to follow this part of the plan.

The testing program:

The following tests were used.

1. The Pre-reading Test to accompany the Sheldon Basic Reading Series Form 1[14] was used to evaluate reading readiness in September 1964. The test was administered in the experimental, control, and sub-control groups by the regular classroom teachers. The test provides a measure of auditory discrimination (rhyming words and initial consonants), visual discrimination of word form, comprehension of material read to the pupil and a perceptual motor tracing test.

2. A letter knowledge test of the alphabet was administered individually to each child in the experimental and control groups in September of 1964. Each child was asked to identify all twenty-six capital letters and all twenty-six lower case letters. The letters were arranged in a randomized order for capital letters and a randomized order for the lower case letters. The letters were shown individually to each child and each child was asked to name the letter if he knew it.

3. The Wechsler Intelligence Scale for Children[15] was administered to each child in the experimental and control groups. Fifty-five were administered in October and five in November.

4. The Gray Oral Reading Test, Form D[16] was administered in December of 1964 to each child in the experimental and control groups. The i/t/a class read from transliterated materials and the control group read from t.o.

5. The word lists for measuring the ability to pronounce words in isolation from the Standard Reading Inventory, Forms A and B, were administered in December of 1964 to the experimental and control groups. Form A[17] was administered in t.o. to both groups; Form B was administered in i/t/a to both groups.

6. Form C of the Gray Oral Reading Test[16] was administered at the end of March 1965 to each child in the experimental and control groups. The children in the experimental group read from transliterated materials and the children in the control group read from t.o.

7. The two word lists from Forms A and B of the Standard Reading Inventory[17] were administered again in March of 1965. Form A was administered in t.o. to each child in the experimental and control groups and Form B in i/t/a to each child in the experimental class and in t.o. to each child in the control group.

8. Form B of the Standard Reading Inventory[17] was administered in its entirety at the end of May 1965 to each child in the experimental class, to each child in the control group and to each child in a group randomly selected from the sub-control group. The test was printed in i/t/a for the experimental group and in t.o. for the control group and the randomly selected group from the sub-control group.

9. The 1964 Stanford Achievement Tests, Form W[18] were administered during the last week of May 1965 to all children in the experimental, control, and sub-control groups. The tests were printed in t.o. for all groups. The complete battery was administered. The tests are word reading, paragraph meaning, vocabulary, spelling, word study skills, and arithmetic. A variation from manual directions was made in administering the spelling test. The children in the experimental class were told to spell in both i/t/a and t.o. if they could, and the pupils were encouraged to attempt to spell the words in t.o. The scoring of the spelling test was done according to t.o. spelling. The Stanford Achievement Tests were administered by the regular classroom teacher.

10. Forms B and D of the Gray Oral Reading Test[16] were administered in

September 1965 to each child in the experimental and control groups. Form B was administered in i/t/a to both groups, and Form D in t.o. to both groups. The purposes of this were to evaluate transfer and to evaluate loss or gain over the summer.

11. Form C of the Gray Oral Reading Test[16] was administered at the end of January of 1966. This was administered in t.o. to both experimental and control groups.

12. Form A of the Standard Reading Inventory[17] was administered at the end of second grade (May 1966) to each child in the experimental class, to each child in the control group, and to each child in the randomly selected sub-control group who was tested at the end of grade one. All the testing was done in t.o.

13. The Stanford Achievement Tests, Form W, Primary II were administered during the last week of May 1966 to all children in the experimental and control and sub-control groups. The tests were administered in t.o. for all groups.

Children were randomly assigned to the order in which they were tested and examiners were randomly assigned to the children in all of the individual testing which was done. Children were assigned randomly according to their placement in the experimental or control groups to two lists. The two lists were used to keep the percentage of pupils tested from the experimental and control groups approximately equal. This seemed important only in the administration of the WISC which took a period of six weeks. Six days was the maximum time lapse in completing the administration of the other individually administered tests. Even so the random assignment was used throughout for both pupil and examiner. Fourteen examiners were administered the individual tests. Four or more examiners participated in each individual testing except the alphabet knowledge test and the WISC.

A time study of pupil activity was conducted throughout grade one in the experimental and control groups. The pupils in both groups were listed randomly with the first pupil listed being the first one observed. Each observer made time studies in both classrooms with the date of the observations in the two classrooms being as close as was possible, except for one observer who made only one observation. This was done so that examiner bias would tend to cancel itself if bias were present. Thirteen observers took part in the time study. All were qualified teachers. The observers were asked to record in minutes the amount of time a child spent during the day in reading, in writing, in phonics, in arithmetic, and in miscellaneous activities. They were asked to divide each category into pupil-work or teacher-work. If the child was in any group large or small and being taught by the teacher the time spent in teaching was recorded under teacher direction in the area taught regardless of the attentiveness of the child. A child was recorded as working individually when he was working independently and individually on a self-instigated project or on a teacher assignment. The observers were instructed to count as individual working time in reading, writing, phonics, and arithmetic only those minutes during which the child was actually working. If a child was given a seat work assignment and then proceeded to sharpen a pencil, get a drink, etc., this time was recorded as miscellaneous pupil-work time. Observations were started twenty minutes before school officially began with the expectation that some children arriving early might choose to participate in individual reading or writing activities. Observations took place approximately every two weeks. A school calendar was obtained at the beginning of the year so that observations could be planned to avoid vacation days but without the foreknowledge by the teachers or principal. Observations were made on all five days of the week.

Sixty time studies were made, thirty in each class. A full listing of the dates and the observers is found in Appendix A.

Instructional programs:

It was planned to begin instruction on the first day of school using the Early-to-Read i/t/a Series in the experimental group and Ginn Basic Reading Series in the control group. The i/t/a materials were late in arriving so that specific instruction with the Early-to-Read i/t/a Series was delayed until the sixth day of school. This was somewhat frustrating to everyone but the delay of six days does not seem significant in evaluating the year's program. The i/t/a teacher was instructed to follow the Early-to-Read Series manuals as closely as possible. She was judged to be successful in doing this, although admittedly such a judgment is subjective.

Beginning books which were available in i/t/a print were purchased for the i/t/a classroom library in addition to the Early-to-Read i/t/a Series. Approximately 160 beginning reading trade books were transliterated specifically for use in the i/t/a classroom. The transliterated books, however, were not available until that time. Each beginning book transliterated for the experimental group was placed in a t.o. edition in the control classroom. Usually two or more copies of each transliterated book were placed in the experimental classroom library and single copies in the control classroom library. This was done because there were more children in the experimental class and because regular library books were not available in i/t/a for the pupils in the experimental class.

The transliterated books were constructed by pasting the transliterated text directly over the traditional print so that the pictures and illustrations were not occluded. The i/t/a material was hinged in place so that

a child could look underneath to practice transfer if he wished. It was noted that most children exhibited no desire at first to look underneath. It was not until they had gained considerable facility in reading that children looked underneath and practiced in t.o.

Statistical tests used:

An analysis of variance was used to analyze the results of the Pre-Reading Test, the WISC, and the Stanford Achievement Tests. A t-test was used to check the significance of differences between groups on the Stanford Achievement Tests, and the grade two results of the SRI.

The Wilcoxon Two Sample Test[19] was used to analyze the results of the alphabet knowledge test, the Gray Oral Reading Paragraph Tests, and the time study. A chi-square test of independence[20] was used to analyze the grade one results of the SRI among the three groups. A sign test[21] was used to analyze the results of the SRI word list tests within the control and experimental classes.

The results:[22]

The results of the Pre-Reading Test indicate that there were no significant differences among the experimental group, the control group and the sub-control group in readiness to read ($F = 1.51$). There was no significant interaction ($F < 1$). There were significant differences among the sub-tests of the Pre-Reading Test ($F = 698.62$; $p < 0.001$), but this would be expected. These results indicate that all three groups were drawn from the same population.

The mean scores for the Pre-Reading Test are reported in Table I. The summary of the analysis of variance for the Pre-Reading Test is reported in Table II.

The mean I.Q. as measured by the Wechsler Intelligence Scale for Children for the experimental group was 106, with a mean Verbal I. Q. of 102 and a mean performance I. Q. of 108. The

mean I.Q. as measured by the WISC for the control group was 102, with a mean verbal I.Q. of 102 and a mean performance I.Q. of 102. There were no significant differences between the two groups ($F < 1$). There was no significant interaction ($F = 1.13$). The mean standard scores for the WISC are reported in Table III. The summary of the analysis of variance for the WISC is reported in Table IV. The mean I.Q. for those children still available for testing at the end of grade two was 103 for both the control and experimental groups.

The median number of lower case letters recognized by the experimental group was 9.0 and by the control group was 10.0. The median number of capital letters recognized by the experimental group was 15.5 and by the control group was 17.5. The Wilcoxon Two Sample Test yielded z-scores of 0.27 and 0.31 for these two differences. Neither is significant. The summary of the Wilcoxon Two Sample Test analysis is reported in Table V.

The results of the three tests given at the beginning of the school year--the Pre-Reading Test, the WISC, and the alphabet knowledge test--indicate no significant differences between the experimental group and the control group. It seems reasonable to conclude that these two classes were drawn from the same population.

The results of the Pre-Reading Test indicate that the sub-control group was also drawn from this same population.

The Wilcoxon Two Sample Test analysis of the time study is reported in Table VI. The results of this analysis indicate that significantly more time was devoted to teacher instruction of the pupils in reading in the control group, and significantly more time was devoted to teacher instruction in arithmetic in the experimental group. Significantly more time was spent

in independent pupil writing activities in the experimental group.

When teacher and pupil times were combined, the control group was found to have spent significantly more time on reading, and the experimental group significantly more time in writing and arithmetic. If the reading and writing times are combined there is no significant difference between the two groups. These differences seem to reflect the differences in the two methods, since the Early-to-Read i/t/a Series uses writing as a basic part of its method of teaching reading.

There were no significant differences noted in the total amounts of time teachers spent in teaching or in the total amounts of time pupils spent in working independently. There were no significant differences noted in the categories labelled phonics or miscellaneous.

TABLE I

RAW SCORE MEANS ON THE PRE-READING TEST ACHIEVED BY THE EXPERIMENTAL GROUP, THE CONTROL GROUP AND THE SUB-CONTROL GROUP IN SEPTEMBER, 1964

	Experimental Group 1	Control Group 2	Sub-Control Group 3
Rhyming	18.5	19.6	18.1
Initial Consonants	16.6	16.8	15.6
Visual Discrimination	18.1	17.7	17.1
Comprehension	16.2	15.0	15.2
Perceptual-Motor	34.6	34.4	33.0

TABLE II

SUMMARY OF ANALYSIS OF VARIANCE ON THE PRE-READING TEST ADMINISTERED IN SEPTEMBER 1964 FOR THE EXPERIMENTAL GROUP, THE CONTROL GROUP, AND THE SUB-CONTROL GROUP

Source	df	ss	ms	F
Subjects	(150)	(8761.82)		
Groups (G)	2	174.88	87.44	1.51
Error b	148	8586.94	58.02	
Within Subjects	(604)	(42193.80)		
Parts (P)	4	34763.20	8690.80	698.62***
GxP	8	64.07	8.01	<1
Error w	592	7366.53	12.44	
Total	754	50955.62		

***$p \leq 0.001$

TABLE III

STANDARD SCORE MEANS ON THE WISC ACHIEVED BY THE EXPERIMENTAL GROUP AND THE CONTROL GROUP IN OCTOBER AND NOVEMBER 1964

	Experimental	Control	Both
Verbal	57.50	56.65	51.13
Performance	60.53	56.69	58.87

TABLE IV

WISC SUMMARY OF ANALYSIS OF VARIANCE FOR THE EXPERIMENTAL GROUP AND CONTROL GROUP

Source	df	ss	ms	F
Subjects	(59)	(10052.000)		
Groups (G)	1	161.560	161.560	<1
Errors b	58	9890.440	170.524	
Within subjects	(60)	(3542.000)		
Parts	1	90.140	90.140	1.54
GxP	1	65.895	65.895	1.13
Error w	58	3385.965	58.378	
Total	119	13594.000		

TABLE V

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE ALPHABET KNOWLEDGE TEST ADMINISTERED TO THE EXPERIMENTAL AND CONTROL GROUPS IN OCTOBER 1964

Test	Medians		z
	experimental	control	
lower case	9.0	10.0	0.27
capitals	15.5	17.5	0.31
N	34	24	

TABLE VI

WILCOXON TWO SAMPLE TEST ANALYSIS OF 30 EXPERIMENTAL GROUP AND 30 CONTROL GROUP TIME STUDIES IN FIRST GRADE DURING THE 1964-65 SCHOOL YEAR

Comparison and categories[a]		Medians (in minutes)			z
		Experimental	Control	Both	
Teacher time in instructing pupils	1	11.33	27.17	26.00	2.61**
	2	2.33	1.83	2.00	1.08
	3	6.00	5.10	5.20	.47
	4	19.67	12.50	16.00	2.81**
	5	27.00	25.00	26.75	.54
	6	74.50	75.25	75.38	.30
Pupil time in working independently	1	19.00	27.25	22.75	1.37
	2	11.00	5.00	8.75	2.34**
	3	.41	4.00	3.33	.76
	4	10.25	12.00	10.00	.97
	5	158.00	153.25	158.00	.56
	6	213.00	215.00	214.63	.45
Teacher and pupil time together	1	45.83	53.00	48.00	2.36**
	2	18.00	8.25	12.00	2.67**
	3	15.00	15.00	15.00	.78
	4	30.00	20.00	25.00	2.65**
	5	174.00	193.50	186.75	1.57
	6	290.00	289.96	289.98	.21
% of each classification spent in teaching	1	49.75	48.50	49.75	.89
	2	20.75	17.50	20.75	.13
	3	51.00	50.00	50.25	.36
	4	65.75	50.00	59.00	1.70*
	5	16.00	14.63	14.80	.32
	6	26.50	25.88	26.08	.32
% of total time spent in each category	1	15.70	18.25	16.63	2.35**
	2	6.00	3.10	4.00	2.75**
	3	5.00	5.00	5.00	.80
	4	10.38	7.00	8.58	2.77**
	5	59.75	69.50	64.13	1.52

(a) 1 - reading; 2 - writing; 3 - phonics; 4 - arithmetic; 5 - miscellaneous; 6 - total

*$p \leq 0.05$
**$p \leq 0.01$

Table VII is the sign test analysis within the experimental and control groups for the SRI word recognition sub-test scores on Forms A and B administered in December and March. The mean scores are reported in Table VIII. The mean number of words recognized in December by the experimental group in reading from t.o. on Form A was 9.67 and the mean from i/t/a on Form B was 21.94. The mean number of words recognized in March by the experimental group in reading from t.o. on Form A was 44.50 and the mean from i/t/a on Form B was 140.40. These differences are highly significant ($p < .001$). The experimental group recognized significantly more words in i/t/a than they did in t.o. This was to be expected.

There was a significant difference in December between the control class scores on Form A and Form B in word recognition. The mean number of words recognized in December by the control group in t.o. on Form A was 14.56 and in i/t/a on Form B was 10.40. This difference is significant at the 0.008 level. The mean number of words recognized in March by the control group in t.o. on Form A was 39.15 and on Form B was 37.81. The difference is not significant. This was expected since both forms were administered in t.o.

Tables IX, X, XI and XII give the chi-square test of independence analyses for these same word pronouncing tests. The median score achieved on Form A in December by the experimental group was 1.64 words pronounced correctly, and 12.20 by the control group. The median score achieved on Form B in December by the experimental group was 4.50 words pronounced correctly, and 9.33 by the control group.

In December the control group scored significantly better than the experimental group on both Form A and Form B word lists. This might not seem so for Form B judging from the means. The scores are not normally

distributed and a median test for differences was applied. The mean score for the i/t/a class was raised well above the median almost entirely by the score of one child who scored 173.

The median number of words pronounced correctly on Form A of the word recognition test in March in reading from t.o. was 36.00 by the experimental group and 37.17 by the control group. The difference is not significant.

The median number of words pronounced correctly on Form B of the word recognition test in March was 140.00 by the experimental group in reading from i/t/a and 36.00 by the control group in reading from t.o. The difference is highly significant ($p < 0.001$).

TABLE VII

SIGN TEST DIFFERENCES WITHIN THE EXPERIMENTAL AND CONTROL GROUPS BETWEEN FORM A AND FORM B SCORES OF THE WORD RECOGNITION SUB-TEST OF THE STANDARD READING INVENTORY ADMINISTERED IN DECEMBER AND MARCH OF FIRST GRADE

	December 1964		March 1965	
Sign (A-B)	Experimental	Control	Experimental	Control
+	4	17	0	15
0	3	3	0	3
-	27	5	32	8
p	< 0.001	0.008	< 0.001	0.105
Administered in	A in t.o. B in i/t/a	A in t.o. B in i/t/a	A in t.o. B in i/t/a	A in t.o. B in t.o.

TABLE VIII

MEAN SCORES OF THE WORD RECOGNITION SUB-TEST OF THE STANDARD READING INVENTORY, FORM A ADMINISTERED IN t.o. AND FORM B ADMINISTERED IN i/t/a IN DECEMBER AND MARCH OF FIRST GRADE

Form	December 1964 experimental	December 1964 control	March 1965 experimental	March 1965 control
A	9.67	14.56	44.50	39.15
B	21.94	10.40	140.40	37.81

TABLE IX

CHI-SQUARE TEST OF INDEPENDENCE AND MEDIAN NUMBER OF WORDS PRONOUNCED CORRECTLY FOR WORD RECOGNITION IN ISOLATION AS MEASURED BY FORM A OF THE STANDARD READING INVENTORY ADMINISTERED IN t.o. TO THE EXPERIMENTAL AND CONTROL GROUPS IN DECEMBER OF FIRST GRADE (1964)

Number of Words Pronounced Correctly	Experimental	Control	X^2
0-1	16	0	
2-5	10	3	31.37***
8-14	1	15	
15 and more	7	7	
median	1.64	12.20	

***$p \leq 0.001$

TABLE X

CHI-SQUARE TEST OF INDEPENDENCE AND MEDIAN NUMBER OF WORDS PRONOUNCED CORRECTLY FOR WORD RECOGNITION IN ISOLATION AS MEASURED BY FORM B OF THE STANDARD READING INVENTORY ADMINISTERED IN i/t/a TO THE EXPERIMENTAL AND CONTROL GROUPS IN DECEMBER OF FIRST GRADE (1964)

Number of Words Pronounced Correctly	Experimental	Control	X^2
0-3	12	2	
4-6	9	5	
7-13	4	11	10.66**
14 and more	9	7	
median	4.50	9.33	

**$p \leq 0.01$

TABLE XI

CHI-SQUARE TEST OF INDEPENDENCE AND MEDIAN NUMBER OF WORDS PRONOUNCED CORRECTLY FOR WORD RECOGNITION IN ISOLATION AS MEASURED BY FORM A OF THE STANDARD READING INVENTORY ADMINISTERED IN t.o. TO THE EXPERIMENTAL AND CONTROL GROUPS IN MARCH OF FIRST GRADE (1965)

Number of words Pronounced Correctly	Experimental	Control	X^2
0-20	10	4	
21-36	6	7	7.09 n.s.
37-46	5	11	
47 and more	10	4	
median	36.00	37.17	

TABLE XII

CHI-SQUARE TEST OF INDEPENDENCE AND MEDIAN NUMBER OF WORDS PRONOUNCED CORRECTLY FOR WORD RECOGNITION IN ISOLATION AS MEASURED BY FORM B OF THE STANDARD READING INVENTORY ADMINISTERED IN i/t/a TO THE EXPERIMENTAL GROUP AND IN t.o. TO THE CONTROL GROUP IN MARCH OF FIRST GRADE (1965)

Number of words Pronounced Correctly	Experimental	Control	X^2
0-34	1	12	
36-67	3	12	
69-144	13	2	16.27***
145 and above	15	0	
median	140.00	36.00	

***$p \leq 0.001$

Tables XIII, XIV, and XV report the chi-square test of independence analyses of the word recognition sub-test of the Standard Reading Inventory administered in t.o. to the control and sub-control groups, and in i/t/a to the experimental class in May at the end of first grade. The median number of words pronounced correctly was 159.0 for the experimental group, 58.5 for the control group, and 38.0 for the sub-control group. The median differences were highly significant ($p < 0.001$). The median differences between the experimental group and both the control group and sub-control group were highly significant ($p < 0.001$). The median difference between the control and sub-control groups was significant at the 0.02 level ($p = 0.0132$).

TABLE XIII

CHI-SQUARE TEST OF INDEPENDENCE AND MEDIAN ACHIEVEMENT FOR WORD RECOGNITION IN ISOLATION AS MEASURED AT THE END OF FIRST GRADE BY THE STANDARD READING INVENTORY FORM B ADMINISTERED IN i/t/a OR t.o. TO THE EXPERIMENTAL, CONTROL AND SUB-CONTROL GROUPS

Number of words Pronounced Correctly	Experimental	Control	Sub-Control	X^2
64 or more	31	11	2	46.66**
0 to 63	2	15	26	
median	159.0	58.5	38.0	

**$p \leq 0.01$

TABLE XIV

CHI-SQUARE TEST OF INDEPENDENCE WITH YATE'S CORRECTION FOR WORD RECOGNITION IN ISOLATION AS MEASURED AT THE END OF FIRST GRADE BY THE STANDARD READING INVENTORY FORM B ADMINISTERED IN i/t/a OR t.o. TO THE EXPERIMENTAL AND CONTROL GROUPS

Number of words Pronounced Correctly	Experimental (n=33)	Control (n=26)	X^2
108 and above	27	3	25.99***
0 to 105	6	23	$p < 0.0001$[a]

a. Based on the Fisher Exact Probability Test[23]

***$p \leq 0.001$

TABLE XV

CHI-SQUARE TEST OF INDEPENDENCE WITH YATE'S CORRECTION FOR WORD RECOGNITION IN ISOLATION AS MEASURED AT THE END OF FIRST GRADE BY THE STANDARD READING INVENTORY FORM B ADMINISTERED IN t.o. TO THE CONTROL AND SUB-CONTROL GROUPS

Number of words Pronounced Correctly	Control	Sub-control	X^2
48 and above	18	9	6.01*
0 to 47	8	19	p = 0.0132[a]

a. based on the Fisher Exact Probability Test[24]

*$p \leq 0.05$

The minimum and maximum instructional reading levels as measured by the Standard Reading Inventory in May at the end of first grade for the three groups are reported in Table XVI. The chi-square test of independence analyses for these instructional levels are reported in Tables XVII and XVIII. The experimental class read from materials printed in i/t/a when taking the SRI and the control and sub-control groups read from materials printed in t.o.

The experimental group had a mean minimum instructional grade level of 1.12 with a range from readiness level to 2^2 reader level. The experimental group had a mean maximum instructional grade level of 2.07 with a range from readiness level to 3^2 reader level.

The control group had a mean minimum instructional grade level of 0.87 with a range from readiness level to 2^2 reader level, and a mean maximum instructional level of 1.01 with a range from readiness level to 3^1 reader level.

The sub-control group had a mean minimum instructional grade level of 0.73 with a range from readiness level to primer reader level, and a mean maximum instructional level of 0.77 with a range from readiness level to primer reader level.

The achievement of boys and girls appears to be equal under i/t/a instruction, and appears to favor girls under control and sub-control conditions. This is not checked statistically because there are insufficient cases and insufficient range of performance.

The differences among minimum instructional levels are significant at the 0.05 level of confidence. This difference is between the experimental group and the other two groups. The difference between the control and sub-control groups in the minimum instructional level is not significant.

The differences among maximum instructional levels are highly significant ($p < 0.001$). The differences among all three groups are significant when judged by their maximum instructional levels. The experimental group scored highest, the control group next highest, and the sub-control group lowest.

TABLE XVI

INSTRUCTIONAL READING LEVELS AS MEASURED IN i/t/a OR t.o. BY THE STANDARD READING INVENTORY AT THE END OF FIRST GRADE FOR THE EXPERIMENTAL, CONTROL AND SUB-CONTROL GROUPS

Grade Level	Book Level Achieved	Experimental				Control				Sub-Control			
		Maximum		Minimum		Maximum		Minimum		Maximum		Minimum	
		boys	girls	boys	girls	boys	girls	boys	girls	boys	girls	boys	girls
3.7	3^2	2	3										
3.2	3^1	1	1				1						
2.7	2^2	5	3	1			0		1				
2.2	2^1	1	4	0	2	1	2		0				
1.7	1^2	2	0	3	2	1	0		1				
1.2	Primer	1	0	3	5	1	5	1	2	3	3	2	3
0.7	Pre-Primer	5	4	9	6	11	4	13	8	11	9	12	8
0.2	Readiness	1	0	2	0	0	0	0	0	2	0	2	1
	N	18	15	18	15	14	12	14	12	16	12	16	12
	Mean Grade Level	2.07		1.12		1.01		0.87		0.77		0.73	

TABLE XVII

CHI-SQUARE TEST OF INDEPENDENCE FOR MINIMUM INSTRUCTIONAL READING LEVELS AS MEASURED AT THE END OF FIRST GRADE BY THE STANDARD READING INVENTORY ADMINISTERED IN i/t/a OR t.o. TO THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS

Book Level	Experimental	Control	Sub-Control	X^2
Pre-primer	17	20	24	7.65*
above pre-primer	16	6	4	

*$p \leq 0.05$

TABLE XVIII

CHI-SQUARE TEST OF INDEPENDENCE FOR MAXIMUM INSTRUCTIONAL READING LEVELS AS MEASURED AT THE END OF FIRST GRADE BY THE STANDARD READING INVENTORY ADMINISTERED IN i/t/a OR t.o TO THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS

Book Level Grouping	Experimental	Control	Sub-Control	X^2
pre-primer and primer	10	14	22	29.86***
1^2 through 2^2	4	8	6	
3^1 and 3^2	19	4	0	

***$p \leq 0.001$

The Wilcoxon analysis of the Gray Oral Reading Test scores obtained in December of first grade are reported in Table XIX. The experimental pupils read from transliterated passages printed in i/t/a for all Gray Oral Reading Tests.

The median total passage score achieved in December by the control group was 0.46. The median total passage score achieved in December by the experimental group was 0.10. The difference between the total passage scores achieved by the control and experimental groups in December is significant ($p<0.05$). Both total passage scores rate below 1.0 grade level. The control group made a median of 4.25 word recognition errors and took a median of 40.25 seconds of reading time on paragraph 1 in December. The experimental group made a median of 14.25 word recognition errors and took a median of 110.0 seconds in reading paragraph 1 in December. The differences between both medians are significant ($p<0.01$).

The control class made a median of 13.0 word recognition errors and took a median of 118.0 seconds in reading paragraph 2 in December. The experimental class made a median of 20.5 word recognition errors and took a median of 180.0 seconds in reading paragraph 2 in December. Neither the word recognition nor time differences are significant ($p>0.05$).

Only six children in the experimental group and five from the control group read well enough to be tested beyond paragraph 2 level. This is an insufficient number of cases to test for significant differences.

The Wilcoxon analysis of the Gray Oral Reading Test scores obtained in March are reported in Table XX. The median total passage score achieved in March by the control group was 10.90. The median total passage score achieved in March by the experimental group was 10.25. There is no significant difference between these two medians ($p>0.05$). Both of these total passage scores rate at

1.0 grade level.

The control class made a median of 0.37 word recognition errors and took a median of 17.5 seconds in reading paragraph 1 in March. The experimental class made a median of 0.69 word recognition errors and took a median of 21.0 seconds in reading paragraph 1 in March. There is no significant difference between the median number of word recognition errors ($p>0.05$). The time differences are significant ($p < 0.01$).

The control class made a median of 1.75 word recognition errors and took a median of 46.0 seconds in reading paragraph 2 in March. The experimental class made a median of 1.90 word recognition errors and took a median of 46.5 seconds in reading paragraph 2 in March. Neither of these differences is significant ($p>0.05$).

The control class made a median of 6.20 word recognition errors and took a median of 90.75 seconds in reading paragraph 3 in March. The experimental class made a median of 3.25 word recognition errors and took a median of 81.75 seconds in reading paragraph 3 in March. The difference between the word recognition error medians is significant ($p<0.01$). There is no significant difference between the median times ($p>0.05$).

The control class made a median of 15.0 word recognition errors and took a median of 176.0 seconds in reading paragraph 4 in March. The experimental class made a median of 5.0 word recognition errors and took a median of 106 seconds in reading paragraph 4 in March. The differences between the median number of word recognition errors is significant ($p<0.01$). The difference between the median number of seconds is significant ($p<0.05$).

Fifteen experimental and only two control pupils read paragraph 5 in March. This is an insufficient number of pupils in the control group for

comparisons to be made. The number of pupils in the experimental group who were able to be tested at paragraph 5 level, however, indicates that the experimental group was able to read better.

TABLE XIX

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE GRAY ORAL READING TEST SCORES ADMINISTERED TO THE EXPERIMENTAL AND CONTROL GROUPS IN DECEMBER 1964

Paragraph Level	Type of Performance	Medians and Number of Pupils			z Score
		Experimental (1)	Control (2)	Both	
1	N	31	25	56	
	Errors	14.25	4.25	7.00	-3.96**
	Time[a]	110.00	40.25	59.00	-4.40**
2	N	18	23	41	
	Errors	20.50	13.00	13.88	-1.18
	Time[a]	180.00	118.00	120.25	-1.58
3	N	6	5	11	
Total passage score		.10	.46	.20	1.96*

a. Time is in seconds; the higher the number the slower the reading.

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XX

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE SCORES OF THE GRAY ORAL READING TEST ADMINISTERED TO THE EXPERIMENTAL AND CONTROL GROUPS IN MARCH 1965

Paragraph level	Type of Performance	Medians and Number of Pupils			z score
		experimental	control	both	
1	N	33	26	59	
	Errors	.69	.37	.48	-1.00
	Time[a]	21.00	17.50	19.38	-2.34*
2	N	32	25	57	
	Errors	1.90	1.75	1.84	- .19
	Time[a]	46.50	46.00	46.00	- .32
3	N	29	21	50	
	Errors	3.25	6.20	4.70	3.08**
	Time[a]	81.75	90.75	83.00	.56
4	N	25	15	40	
	Errors	5.00	15.00	8.17	4.27**
	Time[a]	106.00	176.00	141.50	2.10*
5	N	15	2	17	
Total passage score		10.25	10.90	10.69	.49

a. Time is in seconds; the higher the number, the slower the reading.

*$p \leq 0.05$
**$p \leq 0.01$

Table XXI gives the results of the Stanford Achievement Tests Primary I, form W, administered at the end of first grade to the experimental, control, and sub-control groups. Table XXII gives a summary of the analysis of variance of the results of Stanford Achievement Tests for the three groups. Table XXIII gives the t-test analyses of the differences between groups for the results of the Stanford Achievement Tests.

The analysis of variance indicated significant differences among the groups for all six tests of the SAT. The differences are significant in all comparisons ($p<0.01$).

The mean score for the Word Reading Test was 25.3 for the experimental class, 21.5 for the control class, and 18.6 for the sub-control. The difference in mean score between the experimental and control groups, and the difference in mean score between the experimental and sub-control groups are significant ($p\leq 0.01$). The difference in mean score between the control and sub-control groups is significant ($p=0.021$).

The mean score for the Paragraph Meaning Test was 22.3 for the experimental group, 19.3 for the control group, and 15.3 for the sub-control group. There is no significant difference ($p>0.05$) in mean score between the experimental and sub-control groups and the control and sub-control groups are significant ($p\leq 0.01$).

The mean score for the Vocabulary Test was 26.8 for the experimental class, 25.2 for the control class, and 21.8 for the sub-control class. There is no significant difference ($p>0.05$) in mean score between the experimental and sub-control groups and between the control and sub-control groups are significant ($p<0.01$).

The mean score for the Spelling Test was 9.9 for the experimental group, 11.4 for the control group, and 6.5 for the sub-control group. There is no

significant difference in mean score between the experimental and control groups. The differences in mean score between experimental and sub-control groups and between the control and sub-control groups are significant ($p<0.01$).

The mean score for the Word Study Skills Test was 41.8 for the experimental group, 40.6 for the control group, and 34.0 for the sub-control group. There is no significant difference ($p>0.05$) in mean score between the experimental and control groups. The differences in mean score between the experimental and sub-control group and between the control and sub-control group are significant ($p<0.01$).

The mean score for the Arithmetic Test was 39.6 for the experimental group, 43.7 for the control group, and 34.6 for the sub-control group. There is no significant difference ($p>0.05$) in mean score between the experimental and control groups. The difference in mean score between the experimental and sub-control groups and between the control and sub-control groups are significant ($p\leq0.01$).

The median grade level achieved was 1.95 for the experimental group, 2.00 for the control group, and 1.70 for the sub-control group according to the results of the SAT.

TABLE XXI

RAW SCORE MEAN ACHIEVEMENT AND MEDIAN ACHIEVEMENT ON THE STANFORD ACHIEVEMENT TESTS ADMINISTERED AT THE END OF FIRST GRADE FOR THE EXPERIMENTAL, CONTROL AND SUB-CONTROL GROUPS

Test	Experimental group N=33	Control group N=26	Sub-Control group N=70 & 72
Word Reading	25.3	21.5	18.6
Paragraph Meaning	22.3	19.3	15.3
Vocabulary	26.8	25.2	21.8
Spelling	9.9	11.4	6.5
Word Study Skills	41.8	40.6	34.0
Arithmetic	39.6	43.7	34.6
Median Grade Level Achieved	1.95	2.00	1.70

TABLE XXII

SUMMARY OF ANALYSIS OF VARIANCE FOR STANFORD ACHIEVEMENT TESTS, PRIMARY I, FOR THE EXPERIMENTAL AND CONTROL CLASSES AT THE END OF FIRST GRADE

Test	Source	d.f.	s.s.	m.s.	F
Word Reading	Groups	2	1023.208	511.604	13.32**
	Within groups	126	4840.280	38.414	
	Total	128	5863.488		
Paragraph Meaning	Groups	2	1171.597	585.799	10.43**
	Within Groups	128	7186.449	56.144	
	Total	130	8358.046		
Vocabulary	Groups	2	624.095	312.048	11.52**
	Within groups	128	3467.554	27.090	
	Total	130	4091.649		
Spelling	Groups	2	558.229	279.115	10.42**
	Within groups	128	3429.908	26.796	
	Total	130	3988.137		
Word Study Skills	Groups	2	1726.071	863.036	11.30**
	Within groups	126	9622.735	76.371	
	Total	128	11348.806		
Arithmetic	Groups	2	1705.639	852.820	8.19**
	Within groups	126	13127.865	104.189	
	Total	128	14833.504		

**$p \leq 0.01$

TABLE XXIII

t-TEST ANALYSIS FOR SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF THE THREE GROUPS ON THE STANFORD ACHIEVEMENT TESTS, PRIMARY I, AT THE END OF FIRST GRADE

Test	experimental and control	control and sub-control	experimental and sub-control
Word Reading	2.34**	2.04*	---** [a]
Paragraph Meaning	1.53	2.33**	---**
Vocabulary	1.17	2.86**	---**
Spelling	1.11	---**	3.13**
Word Study Skills	0.52	3.29**	---**
Arithmetic	1.53	---**	2.32**

a. t-scores were not computed when the difference between means was greater than the already computed significant difference between the other two means.

*$p \leq 0.05$

**$p \leq 0.01$

The Wilcoxon Analysis of the Gray Oral Reading Test scores obtained in September of second grade are reported in Tables XXIV, XXV, XXVI and XXVII. Both the control and experimental pupils took two forms of the test. Form B was administered in i/t/a and Form D was administered in t.o.

There was no significant difference between the experimental and control groups in the total score achieved on Form D administered in t.o. ($p=0.52$). At paragraph level 4 there was a significant difference in errors made ($p<0.01$) and the time of reading ($p<0.05$). The performance of the i/t/a group was better.

There was no significant difference ($p>0.05$) between the performance of the experimental pupils when reading from i/t/a or t.o. This would indicate that transfer had been obtained; however, this seemed to be a regression for the experimental group from their June i/t/a reading rather than a gain in performance when reading from t.o.

There were significant differences when the t.o. taught children were asked to read in i/t/a. The t.o. children were significantly poorer ($p<0.01$) than the i/t/a children when reading from i/t/a, and t.o. children were significantly poorer ($p<0.01$) when reading from i/t/a than they were when reading from t.o. This would have been predicted. It is interesting to note that five or six of the t.o. taught children were able to transfer without loss to i/t/a in the September reading. These were pupils who were reading well in t.o.

The Wilcoxon Analysis of the Gray Oral Reading Test scores, Form C, obtained in January at the middle of second grade are reported in Table XXVIII. This was administered in t.o. There are no significant differences in total performance or on any of the subtest scores achieved.

Table XXIX gives the results of the Stanford Achievement Test Primary II,

Form W administered at the end of second grade to the experimental, control, and sub-control groups. The i/t/a taught pupils achieved the highest average subtest scores on all eight tests. The median grade level achieved on the SAT was 3.20 for the experimental group, 2.70 for the control group, and 2.55 for the sub-control group.

Table XXX gives a summary of the Analysis and Variance of results of the Stanford Achievement Tests for the three groups. The analysis of variance indicated significant differences ($p<0.01$) among the groups for six of the eight tests of the Stanford Achievement Tests. There were significant differences on the test of word meaning, paragraph meaning, word study skills, language, arithmetic computation, and arithmetic concepts. There were no significant differences ($p>0.05$) in science and social studies concepts or spelling.

Table XXXI gives a t-Test Analysis of the differences between groups for the results of the Stanford Achievement Tests at the end of grade two. There were no significant differences ($p>0.05$) between the scores of the experimental and control group except in Arithmetic Concepts ($p<0.01$). There were significant differences between the experimental and sub-control group in word meaning ($p<0.05, >0.01$), paragraph meaning ($p<0.01$), word study skill ($p<0.01$), language ($p<0.05, >0.01$), and arithmetic concepts ($p<0.01$). There were no significant differences ($p>0.05$) between the scores achieved by the experimental and sub-control groups in science and social studies concepts, spelling, and arithmetic computation.

There were no significant differences ($p>0.05$) between the scores achieved by the control and sub-control groups on the eight tests of the SAT administered at the end of grade two.

The instructional reading levels as measured by the Standard Reading Inventory at the end of second grade are reported in Table XXXII. The experimental group achieved a maximum instructional reading level of 3.1 which would place them as ready to begin the 3-1 basal book. The control group achieved a maximum instructional level of 2.5 which would place them as ready to begin the 2-2 reader. The sub-control group achieved a maximum instructional grade level of 2.0 which would place them as ready to begin the 2-1 book. The minimum instructional level achieved by the experimental pupils was 2.0. The minimum instructional level achieved by the control pupils was 1.5 and the minimum instructional level achieved by the sub-control group was 1.3.

Table XXXIV reports the t-test analysis for significance of differences between the means of the experimental, control and sub-control groups on the Standard Reading Inventory at end of grade two. There were significant differences ($p \leq 0.01$) between the experimental and sub-control groups on both maximum and minimum instructional levels. There were no significant differences ($p > 0.05$) between the experimental and control group or between the control and sub-control group on maximum and minimum instructional levels.

Table XXXIII reports the mean achievement scores on the Standard Reading Inventory subtests and the t-test comparison between the experimental, control, and sub-control groups. There are nine subtests on the Standard Reading Inventory. These subtests are (1) pronouncing vocabulary in context, (2) pronouncing vocabulary in isolation, (3) word recognition errors in oral reading, (4) total errors in oral reading, (5) comprehension-recall after oral reading, (6) comprehension-recall after silent reading, (7) comprehension-total interpretation after oral and silent reading, (8) speed of oral

reading, and (9) speed of silent reading.

The pronouncing vocabulary in isolation test on the Standard Reading Inventory is administered separately from the other tests. Pupils are asked to pronounce from words presented in isolation. There are 265 words in the complete test beginning at pre-primer and running through seventh reader level. The mean score achieved by the experimental pupils was 181.45, by the control pupils 130.89, by the sub-control pupils 98.52. The difference between the experimental and control, and the difference between experimental and sub-control were significant ($p<0.01$), and the difference between the control and sub-control was significant ($p<0.05, >0.01$). The mean achievement for these same experimental group pupils at the end of grade one was 175.14. Fourteen pupils scored better at the end of grade two and fifteen scored better at the end of grade one. The change in scores is not significant ($p>0.05$). The mean achievement for these same control group pupils at the end of grade one was 60.94. All eighteen pupils scored better at the end of grade two. The change in scores is significant ($p<0.001$).

In all of the other SRI subtest measures there was only one significant difference. The experimental group was significantly better ($p<0.01$) than the sub-control group on the pronouncing vocabulary in context subtest. On the SRI subtests there was no consistent superiority of one group over the others. The Standard Reading Inventory subtest scores, except for the pronouncing vocabulary in context and the pronouncing vocabulary in isolation, are scores achieved in reading from preprimer, primer, and first reader materials only. Some of the children in each group were so frustrated by levels at 2-1 reader level and above that it was impossible to test at these levels and, therefore, impossible to make comparisons of subtest scores above 1-2 level since such comparisons would be throwing out the poorest readers of each group in

unequal proportions.

The types of oral reading errors made when reading from the Standard Reading Inventory were tabulated into eight error categories:

1. Pronunciation. This error occurred when the examiner had to pronounce a word for the child.
2. Mispronunciation. This error occurred when a child attempted to pronounce a word and distinctly mispronounced it.
3. Mumbling. This error occurred when a child mumbled a word in such a manner that the examiner could not understand it or record it phonetically.
4. Repetition. This error occurred when a child repeated a syllable, word, or phrase.
5. Substitution. This error occurred when a child substituted one word for another.
6. Omission. This error occurred when a child omitted a part of a word, a word, or a phrase.
7. Addition. This error occurred when a child added an ending, a word, or a phrase.
8. Punctuation. This error occurred whenever a child definitely misread the punctuation.

The number of oral reading errors made by the three groups, the percentages of each type of error made, and the chi-square comparisons of the error tabulations are recorded in Table XXXV. There were significant differences in the error patterns. The sub-control group had a different error pattern than either the i/t/a or t.o. group. The i/t/a and main control groups were not different in their error patterns. The sub-control group needed much more help in having words pronounced for them by the examiner and tended to

wait for assistance, thereby making fewer repetitions. The differences found reflect primarily the differences in level of achievement. Children who are achieving at lower levels or beginning reading levels tend to want more assistance from a teacher in pronouncing words and have been found to have a higher percentage of pronunciation errors. As children gain a mastery of word decoding skills, there is a tendency for less pronunciation help and for more repetition as a child repeats in decoding words and phrases.

There is one difference on the testing between the i/t/a taught children and the control and sub-control groups which was noted consistently throughout the two years which is not discernible from the test scores achieved. The i/t/a children consistently attempted more paragraphs or stories when reading from the Gray Oral Reading Tests and from the Standard Reading Inventory even though they did not achieve better. They seemed to have developed a greater independence or a greater tolerance of frustration. Another way to interpret this would be that they had not been taught to depend upon the teacher for assistance in decoding words. This difference is reflected in the superiority on the Standard Reading Inventory subtest pronouncing words in isolation.

The i/t/a children consistently did well on all of the reading test measures. It should be noted, however, that four children in the i/t/a class still were unable to achieve beyond the preprimer level at the end of second grade and needed further instruction at this level. There were two control children and four sub-control children in this same category. As a group the i/t/a taught children learned to read well, but the use of i/t/a did not eliminate the problem of a child who seems unable to achieve more than marginally in reading in first and second grade.

It should be noted also that the Hawthorne effect, if any, probably was

eliminated from the control children in second grade by the nature of their assignment as a group to one teacher. Random assignment of pupils to the two second grades was not possible. The i/t/a children in second grade still needed special attention in materials new to the teacher, and, since many of the pupils were still writing in i/t/a, there was probably an effect upon the teacher. However, there was a growth in reading in second grade of 1.5 years in the control group, 1.0 year in the experimental group, and 1.3 years in the sub-control group as measured by the maximum instructional level of the Standard Reading Inventory and a growth of 0.6 year in the control group, of 0.9 year in the experimental group, and 0.6 year in the sub-control group as measured by the minimum instructional level of the S.R.I. The differences between the gains of the control and sub-control groups at maximum and minimum levels (1.5 vs 0.6 and 1.3 vs 0.6) seem marked compared to the lack of difference (1.0 vs 0.9) for the experimental group. This seems to reflect the independence of attack or the willingness to tolerate frustration which was noted previously in the experimental group.

TABLE XXIV

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE GRAY ORAL READING TEST SCORES FOR FORM B ADMINISTERED IN i/t/a TO THE EXPERIMENTAL AND CONTROL GROUPS AT THE BEGINNING OF SECOND GRADE
September 1965

Paragraph Level	Number and type of performance	Medians and Numbers of Pupils		z score
		experimental	control	
1	N	29	22	
	Errors	1	5	4.26**
	Time	17	55	4.83**
2	N	28	20	
	Errors	2	9	4.66**
	Time	35	104	4.40**
3	N	24	6	
	Errors	4	4	0.65
	Time	46	72	1.43
4	N	21	3	
	Errors	3	5	2.01*
	Time	52	80	1.53
5	N	18	2	
	Errors	6	8	0.31
	Time	76	73	0.44
Total passage	N	30	22	
		2	0	4.67**

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XXV

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE GRAY ORAL READING TEST SCORES FOR FORM D ADMINISTERED IN t.o. TO THE EXPERIMENTAL AND CONTROL GROUPS AT THE BEGINNING OF SECOND GRADE
September 1965

Paragraph level	Number and type of performance	Medians and Numbers of Pupils		z score
		experimental	control	
1	N	29	22	
	Errors	1	1	-0.98
	Time	18	19	0.08
2	N	28	22	
	Errors	2	2	-0.33
	Time	31	39	0.39
3	N	25	20	
	Errors	3	4	0.79
	Time	48	67	1.40
4	N	20	11	
	Errors	4	10	2.75**
	Time	60	107	2.23*
5	N	15	5	
	Errors	5	10	0.61
	Time	64	101	1.09
6	N	8	3	
	Errors	7	12	1.33
	Time	72	88	0.71
Total passage	N	30	22	
		0	9	-0.52

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XXVI

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE GRAY ORAL READING TEST SCORES FOR FORM D ADMINISTERED IN t.o. AND FORM B ADMINISTERED IN i/t/a TO THE EXPERIMENTAL GROUP AT THE BEGINNING OF SECOND GRADE, SEPTEMBER 1965

Paragraph level	Number and type of performance	Medians and Numbers of Pupils		z score
		Form B	Form D	
1	N	27	28	
	Errors	1	1	0.47
	Time	17	18	0.11
2	N	28	28	
	Errors	2	2	0.49
	Time	35	31	0.02
3	N	24	25	
	Errors	4	3	0.46
	Time	46	48	0.23
4	N	20	20	
	Errors	3	4	0.45
	Time	52	60	0.51
5	N	17	15	
	Errors	8	5	1.08
	Time	78	64	1.42
6	N	11	8	
	Errors	11	7	-0.61
	Time	103	72	1.20
Total passage	N	30	30	
		2	0	0.22

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XXVII

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE GRAY ORAL READING TEST SCORES FOR FORM D ADMINISTERED IN t.o. AND FORM B ADMINISTERED IN i/t/a TO THE CONTROL GROUP AT THE BEGINNING OF SECOND GRADE, SEPTEMBER 1965

Paragraph level	Number and type of performance	Medians and Numbers of Pupils		z score
		Form B	Form D	
1	N	22	22	
	Errors	5	1	4.64**
	Time	55	19	5.27**
2	N	20	22	
	Errors	9	2	4.33**
	Time	104	39	4.34**
3	N	6	20	
	Errors	4	4	0.18
	Time	72	67	0.61
4	N	3	11	
	Errors	4	10	-1.16
	Time	80	107	-0.92
5	N	2	5	
	Errors	0	10	-0.38
	Time	73	101	-0.18
Total passage	N	22	22	
		0	9	4.86**

*$p \leq 0.05$

**$p \leq 0.01$

TABLE XXVIII

WILCOXON TWO SAMPLE TEST ANALYSIS OF THE GRAY ORAL READING TEST SCORES FOR FORM C ADMINISTERED IN t.o. TO THE EXPERIMENTAL AND CONTROL GROUPS IN JANUARY 1966 (MID-SECOND GRADE)

Paragraph level	Number and type of performance	Medians and Numbers and Pupils		z score
		i/t/a	t.o.	
1	N	28	19	
	Errors	1	1	-0.29
	Time	12	11	-0.04
2	N	26	19	
	Errors	1	1	0.06
	Time	22	22	1.07
3	N	25	18	
	Errors	1	1.5	0.50
	Time	23	26.5	1.03
4	N	24	16	
	Errors	2	4.5	1.73
	Time	34	47.5	1.71
5	N	18	8	
	Errors	3.5	3.0	-0.49
	Time	39.5	48.5	0.42
6	N	12	6	
	Errors	4	13.5	1.78
	Time	50.0	113.5	1.78
Tot[illegible] passage	N	28	19	
		21.5	20	-0.20

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XXIX

RAW SCORE MEAN ACHIEVEMENT OF THE STANFORD ACHIEVEMENT TEST, PRIMARY II, AT THE END OF SECOND GRADE FOR THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS

Test	Experimental group N=27 & 28	Control group N=17 & 16	Sub-control group N=49-52
Word Reading	22.4	17.7	15.2
Paragraph Meaning	37.5	30.7	25.7
Science and Social Studies Concepts	21.1	18.9	19.0
Spelling	11.3	10.1	8.2
Word Study Skills	41.6	35.2	31.1
Language	40.0	35.3	33.3
Arithmetic Computation	23.5	20.6	17.4
Arithmetic Concepts	21.9	15.2	15.6
Median Grade Level Achieved	3.20	2.70	2.55

TABLE XXX

SUMMARY OF ANALYSIS OF VARIANCE OF THE STANFORD ACHIEVEMENT TESTS, PRIMARY II, FOR THE EXPERIMENTAL CONTROL AND SUB-CONTROL GROUPS AT THE END OF SECOND GRADE

Test	Source	d.f.	s.s.	m.s.	F
Word Meaning	Groups	2	916.864	458.432	9.233**
	Within groups	93	4617.136	49.646	
	Total	95	5534.000		
Paragraph Meaning	Groups	2	2483.910	1241.955	9.447**
	Within groups	92	12094.580	131.462	
	Total	94	14578.490		
Science and Social Studies Concepts	Groups	2	83.680	41.840	1.379
	Within groups	90	2728.794	30.319	
	Total	92	2812.474		
Spelling	Groups	2	178.327	89.163	2.325
	Within groups	85	3259.662	38.348	
	Total	87	3437.989		
Word Study Skills	Groups	2	1963.710	981.855	7.902**
	Within groups	92	11430.880	124.248	
	Total	94	13394.590		
Language	Groups	2	816.640	408.320	6.589**
	Within groups	93	5762.520	68.423	
	Total	95	6579.160		
Arithmetic Computation	Groups	2	679.999	339.999	4.969**
	Within groups	93	6363.408	68.423	
	Total	95	7043.407		
Arithmetic Concepts	Groups	2	824.984	412.492	7.616**
	Within groups	92	4982.238	54.154	
	Total	94	5807.222		

**$p \leq 0.01$

TABLE XXXI

t-TEST ANALYSIS FOR SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF THE ACHIEVED BY THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS ON THE STANFORD ACHIEVEMENT TESTS, PRIMARY II, AT THE END OF SECOND GRADE

Test	experimental and control	control and sub-control	experimental and sub-control
Word Meaning	1.99	0.75	2.30*
Paragraph Meaning	1.98	1.59	4.44**
Science and Social Studies Concepts	1.31	0.04	1.15
Spelling	0.59	1.25	1.89
Word Study Skill	1.72	1.35	2.68**
Language	1.85	0.94	2.46*
Arithmetic Computation	0.95	1.28	1.94
Arithmetic Concepts	3.11**	0.18	2.74**

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XXXII

INSTRUCTIONAL READING LEVELS AS MEASURED IN t.o. BY THE STANDARD READING INVENTORY AT THE END OF SECOND GRADE FOR THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS

Grade Level	Book Level Achieved	Experimental				Control				Sub-Control			
		Maximum		Minimum		Maximum		Minimum		Maximum		Minimum	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
5.5	5		1										
4.5	4	1	1										
3.7	3^2	9	6		3	2	3				1		
3.2	3^1	2	1		2	1	3	1	1	1	1		1
2.7	2^2		2	2	2			1	1	4	2		1
2.2	2^1	1	1	5	1	3	1	1		2	1		
1.7	1^2	1		3	2	1			2		2	4	1
1.2	Primer			3	2	2		3	1	2	1	2	3
.7	PrePrimer	3	1	4	1	1	1	4	3	3	1	6	3
.2	Readiness												
	N	17	13	17	13	10	8	10	8	12	9	12	9
	Mean Grade Level	3.1		2.0		2.5		1.5		2.0		1.3	

TABLE XXXIII

MEAN ACHIEVEMENT SCORES ON STANDARD READING INVENTORY SUB-TESTS[1] AND t-TEST COMPARISONS BETWEEN THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS AT THE END OF THE SECOND GRADE

Sub-test	Means					
	experimental N=30	control N=18	sub-control N=21	experimental and control	control and sub-control	experimental sub-control
Pronouncing vocabulary in context[2]	72.13	68.77	59.15	0.688	1.805	2.827**
Pronouncing vocabulary in isolation[2]	181.45	130.89	98.52	3.14**	2.191*	5.413**
Word recognition errors in oral reading[3]	4.31	4.33	3.90	0.014	0.255	0.250
Total errors in oral reading[3]	8.76	10.11	9.10	0.578	0.464	0.330
Comprehension: recall after oral reading[2]	20.50	21.00	18.90	0.656	1.245	1.169
Comprehension: recall after silent reading[2]	13.66	15.41	13.10	1.621	1.445	0.385
Comprehension: total interpretation after oral and silent reading[2]	8.83	9.00	7.65	0.119	0.719	0.954
Speed of oral reading[3]	107.89	134.72	121.30	1.170	0.452	0.617
Speed of silent reading[3]	91.10	77.66	79.90	0.625	0.817	0.503

1. The pronouncing vocabulary in context sub-test covers levels pre-primer through 3-2 reader; the pronouncing vocabulary in isolation sub-test covers levels pre-primer through seventh reader; the other seven sub-tests cover levels pre-primer through 1-2 reader only.
2. A high score is better achievement than a low score.
3. A low score is better achievement than a high score.

*$p \leq 0.05$
**$p \leq 0.01$

TABLE XXXIV

t-TEST ANALYSIS FOR SIGNIFICANCE OF DIFFERENCES BETWEEN THE MEANS OF THE EXPERIMENTAL, CONTROL, AND SUB-CONTROL GROUPS ON THE STANDARD READING INVENTORY AT THE END OF GRADE TWO

Level	experimental and control	control and sub-control	experimental and sub-control
Maximum Instructional	.734	1.658	3.563**
Minimum Instructional	1.805	0.779	2.893**

**$p \leq 0.01$

TABLE XXXV

ORAL READING ERRORS MADE BY THE EXPERIMENTAL (I/T/A), THE CONTROL, (T.O.), AND THE SUB-CONTROL GROUPS WHEN READING ORALLY FROM THE STANDARD INVENTORY IN MAY AT THE END OF THE SECOND GRADE AND THE CHI-SQUARE COMPARISONS

Type of Error	Number of Errors				Percentage of each type				Chi-square		
	Total N=69	i/t/a N=30	t.o N=18	Sub-con. N=21	Total	i/t/a	t.o.	Sub-con.	i/t/a	t.o.	Sub-con.
Pronunciation	300	94	62	144	12	8	10	20	13.08**	2.95	38.22**
Mispronunciation	102	46	44	12	4	4	7	2	0.01	12.28*	10.38**
Mumbling	29	11	2	16	1	1	0.3	2	0.37	4.13	6.88
Repetition	667	335	195	137	27	30	31	19	3.40	3.25	16.00**
Substitution	498	210	123	165	20	19	19	23	1.74	0.51	3.16
Omission	545	272	129	144	22	24	20	20	2.38	0.88	1.1
Addition	315	148	77	90	13	13	12	13	0.18	0.19	0.07
Punctuation	22	10	5	7	1	1	0.7	1	0.00	0.09	0.06
TOTAL	2478	1126	637	715	100	100	100	100	21.15	24.28*	75.87**

*$p \leq 0.05$
**$p \leq 0.01$

One sub-study was added to the main study. The 1964-65 control and experimental first grade teachers again taught first grade during the 1965-66 school year. The experimental teacher continued with the Early-to-Read i/t/a Series and the control teacher continued in t.o. with the Ginn Basic Readers. The children entering first grade were placed in either of the two first grades according to the judgment of the kindergarten teacher. This procedure was customary, but was not random placement as used the preceding year.

The main concern was whether the classes of the two teachers would maintain their 1964-65 achievement levels as assessed by the Stanford Achievement Tests, Primary I, Form W. The SAT were administered to the classes of the control and experimental teachers in May of 1966. The results are in Table XXXVI. The 1965-66 experimental class had a median grade level performance of 1.85. The 1964-65 experimental class had scored a median of 1.95. The mean scores achieved by the 1965-66 experimental class were slightly lower in five of the six tests. Spelling was slightly higher. There were no significant differences in achievement between the 1964-65 and 1965-66 experimental classes.

The 1965-66 control class had a median grade level performance of 2.25. The 1964-65 class had scored a median of 2.00. The mean scores achieved by the 1965-66 control group were higher in all six tests. The 1965-66 control class achieved significantly better in Word Reading ($p<0.01$), Paragraph Meaning ($p<0.05$), and Spelling ($p<0.05$).

The 1965-66 control class achieved significantly better than the 1965-66 experimental class in Word Reading ($p<0.01$), Spelling ($p<0.01$), and Arithmetic ($p<0.01$). This reversed the one significant difference noted

between the 1964-65 groups where the experimental group was significantly better ($p<0.05$) than the control group in Word Reading achievement. There was no significant difference in Spelling achievement between the 1964-65 groups. This lack of difference in 1964-65 is contrary to most other reported findings for achievement the end of the first grade. The difference favoring the 1965-66 control group in spelling is consistent with the findings of other studies of spelling achievement for the end of first grade. The better achievement in arithmetic in the 1965-66 control class is consistent with the direction of the achievement of the 1964-65 study. This superior arithmetic achievement is taken to indicate that the control teacher did not sacrifice her arithmetic program in 1965-66 in order to give more time to reading.

The achievement scores for both years might be interpreted as a negative Hawthorne effect on the control teacher, or the combination of a negative Hawthorne effect during the 1964-65 year and a positive effect during the 1965-66 year. In 1964-65 the control teacher, although given as much observation attention as the experimental teacher, did not receive the community, parental, and school attention that the experimental teacher received. Neighboring school districts asked to visit the i/t/a class. None asked to visit the control class. Some visitors were bootlegged in although visits by outsiders were forbidden under the experimental design. Parents of the i/t/a children were highly interested. Normal interest, only, was displayed by the parents of the control children. The press featured the i/t/a class in an article and editorial, excluding the control class and the experiment. It was felt that the control teacher became increasingly dissatisfied as the year progressed, and in June of 1965, prior to knowledge of results of any of the testing, she had decided that she would shift to i/t/a the next year.

The control teacher changed her mind when she learned that there were generally no significant differences in achievement. She approached the 1965-66 year with the attitude that she would prove that a t.o. basal approach worked well. She had noted the amount of independent writing which had been done in the i/t/a classroom and decided that a t.o. classroom could do as much. She added the S.R.A. Reading Laboratory I: Word Games for Grade one to her program. This supports a positive Hawthorne effect on the control teacher the second year.

TABLE XXXVI

MEAN SCORES ACHIEVED BY 1965-66 FIRST GRADE CLASSES OF THE EXPERIMENTAL AND CONTROL TEACHERS OF THE 1964-65 STUDY AS MEASURED BY THE STANFORD ACHIEVEMENT TEST, PRIMARY I, FORM W, AND THE t-TEST COMPARISONS BETWEEN THE 1965-66 CONTROL AND EXPERIMENTAL GROUPS, THE 1964-65 AND THE 1965-66 EXPERIMENTAL GROUPS, AND THE 1964-65 AND THE 1965-66 CONTROL GROUPS.

Test	MEAN Experimental 1965-66	Control 1965-66	t Experimental vs. Control 1965-66	t Exp. 1964-65 Exp. 1965-66	t Control 1964-65 Control 1965-66
Word Reading	23.32	28.17	2.92**	1.18	4.47**
Paragraph Meaning	21.29	24.17	1.18	0.42	2.02*
Vocabulary	24.96	27.72	1.67	1.30	1.56
Spelling	10.07	14.93	4.05**	0.15	2.24*
Word Study Skills	40.13	41.52	0.59	0.82	0.37
Arithmetic	35.68	45.48	3.50**	1.26	0.63
Median Grade Level Achieved	1.85	2.25			

* $p \leq 0.05$
** $p \leq 0.01$

Conclusions and discussion:

1. Over-all, the amount of time in first grade spent in the teaching of reading and writing did not seem to be different under the Early-to-Read i/t/a Program in the experimental group from that under the traditional program in the control group.

2. There is no significant superiority in over-all reading achievement between the experimental and main control group at the end of first or second grade when the reading is done in t.o. There are consistent, significant differences with better performance by the experimental group in tests requiring only word pronunciation skill.

3. The results of the spelling tests indicate that the Early-to-Read i/t/a Program is not detrimental to t.o. spelling achievement at the end of first or second grade. Pupils taught in i/t/a can be expected to spell in t.o. as well as traditionally taught pupils.

4. The pupils taught under the experimental program had a greater range of achievement and their scores had a more normal distribution than did the others, particularly at the end of grade one when measured by individually administered reading tests in i/t/a. The average and above average pupils seemed to be extended in their achievement. However, a slow beginning achievement for some children was not eliminated.

5. Word pronouncing achievement as measured by the SRI Vocabulary in Isolation Sub-Test was significantly better for the experimental group than for the other groups when reading from i/t/a at the end of grade one and significantly better when reading from t.o. at the end of grade two.

6. Word pronouncing skill as measured by the SRI Vocabulary in Isolation Sub-Test indicates that for all children in all groups word pronouncing achievement is superior

to general ability to read. Children were frequently rated at frustration level in reading at a book level at which they could pronounce correctly all of the vocabulary in isolation. This is not meant to say that these children did not understand the words which they pronounced. They recognized (pronounced and understood) the individual words. They lacked fluency in reading sentences consecutively, their reading was slow, and their oral reading characterized by numerous pauses, poor-phrasing, repetition, mis-called words, etc. causing their performances to be rated at frustration level. This superiority of word pronouncing to overall reading was exhibited by the experimental group at the end of grade one and all groups at the end of grade two.

7. The amount of loss which might be expected in transfer from i/t/a to t.o. is crucial in evaluating this study. One indication of the loss is given on the vocabulary in isolation sub-test of the SRI administered in March of first grade in which the experimental and control groups were not significantly different in their abilities to pronounce words in t.o. The median in March of grade one in pronouncing vocabulary in isolation was 36.00 for the experimental group and 37.17 for the control group in t.o. The median for the experimental group when reading from i/t/a was 140.00 which is significantly greater ($p<0.01$) than the control group and the t.o. performance of the experimental group compared to itself. The median in May of grade one for the experimental group was 159.00 and 58.50 for the control group. The magnitude of both shifts would indicate that the experimental and control groups would not have been significantly different if a t.o. measurement had been made for both groups in May of grade one.

The mean scores of the SRI pronouncing vocabulary in isolation test were 181.44 for the experimental group, 130.44 for the control group, and

98.52 for the subcontrol group at the end of grade two. The mean scores were 175.14 for the same experimental children and 60.94 for the same control children at the end of grade one.

The experimental group did not change significantly ($p=.50$) in its performance at the end of grade one to the end of grade two. In grade one the test was administered in i/t/a, and in t.o. in grade two. This would indicate that children taught with i/t/a do learn to decode very rapidly but that once this skill is mastered we should not expect continued rapid growth. The control group and the subcontrol group achieved significantly better ($p<0.001$) at the end of second grade than they did at the end of first. Both of their test measures were made in t.o. This could indicate that it took a year of work in reading to transfer in this one area. It might mean that something of a maximum had been reached and that transfer was reached in a relatively short time with a plateau of achievement being maintained for several months.

The results of the Gray Oral Reading Test administered at the beginning of grade two showed no significant differences in achievement between the experimental group's scores when tested in i/t/a and t.o. This would indicate a substantial loss in i/t/a achievement over the summer months, since achievement in i/t/a had been superior to achievement in t.o. in the June testing.

The results of the Gray Oral Reading Test administered in the middle of second grade indicate no significant difference in achievement between the experimental and control groups. This would indicate that transfer was not taking place immediately but rather slowly.

The amount of loss in general reading ability on the SRI in shifting from i/t/a to t.o. has not been measured. However, the 1964 study[25] of transfer from t.o. to i/t/a would suggest a minimum loss of 0.5 year, a

probable loss of 1.0 year, and a maximum loss of 1.5 years in transferring. This would suggest that no significant difference in reading achievement would have been found on the SRI if it had been administered in t.o. to the experimental group in May of first grade. The results of the Gray Oral Reading Test at the beginning and middle of second grade support this estimate of loss.

It would seem that the early high achievement in i/t/a is lost in transfer except for skill in pronouncing words in isolation. It would appear that the control pupils are catching up in this area. This result should be neither surprising nor unexpected. Early achievement in reading has not been demonstrated to mean greater achievement later. The problem of learning to read is largely one of word pronouncing at the very beginning stages; thus, i/t/a succeeds well at the beginning stages. However, very few people, if any, equate word pronouncing with reading. Reading is more than word pronouncing. The results of this study support this contention. It seems that transfer does take place, that it takes place without stress, and that it takes a long period of practice for recovery of fluency before growth continues. It would seem unwise to expect that a child will transfer and continue to grow in reading achievement without a period of practice. This period of practice is rather lengthy, averaging a minimum of six months perhaps.

8. The superior achievement of the experimental and control groups as compared to the sub-control group indicates good achievement under both the Early-to-Read i/t/a Program and the traditional program. There are no marked detrimental effects noted for the children in the experimental group. There is one marked superiority, word pronouncing in isolation. This superiority may not be maintained. The results of this study are not clear on this since the experimental pupils seem to have reached a plateau while the control pupils continue to grow.

The possible Hawthorne effect may be one of the more significant findings of this study, although studying the Hawthorne effect was not the purpose or intent of the study. However, the sub-control group was used to check against just such an effect. The results are interesting to compare with the findings of the Bleismer[26] study of ten methods in first grade reading instruction. Bleismer used the SAT Form W to measure achievement at the end of first grade. This is the same test and form used in this study. One group in the Bleismer study used the Ginn Basic Readers. The control group in this study used the same Ginn Basic Readers. The Ginn Reader taught group in Bleismer's study achieved poorly (a median mean grade level of 1.62) and were significantly poorer in achievement than the best four methods. None of the four best methods reported in the Bleismer study would appear to have produced achievement superior to either the 1964-65 or the 1965-66 control group's achievement in this study. These findings could be interpreted as additional evidence of a Hawthorne effect.

The results achieved by the control group in 1964-65 and again in 1965-66 compare favorably with the results of the U.S. Office of Education sponsored studies, although the U.S. Office sponsored studies have 15 to 20 less teaching days in their measurements. The results of the sub-control classes compare unfavorably with the same studies. Again, the evidence supports a Hawthorne effect.

9. Another interpretation, more positive than Hawthorne effect, can be made. Both the control and experimental teacher were stimulated to teach well in the 1964-65 year. The control teacher was stimulated particularly to teach well in 1965-66. Good teaching under traditional orthography appears to be more the answer to the problem of first grade instruction than a change of the alphabet. Our problem is stimulating good teaching, because

the evidence in this study indicates that it can be stimulated. I/t/a can be the stimulant, but it is suspected that i/t/a with unstimulated teachers could produce poor readers as easily as t.o. methods under similar circumstances.

The control and experimental teachers both felt that they had worked harder under experimental conditions than they normally did. Both learned that children can learn to write and write fairly well if they are not hampered by artificial barriers, the teacher's emphasizing handwriting craftsmanship to the detriment of expression, the teacher's correcting spelling to the point that children do not wish to try, the teacher's assuming that children cannot learn vowel sounds until second grade and thereby assuming that most independent word attack skills must wait until second grade, and the teacher's emphasizing the child's dependence upon the teacher for help in learning to read, write, and spell. Both teachers became aware of the importance of auditory discrimination, and both acknowledged that they learned a great deal about auditory discrimination which they had not known before.

Summary:

Thirty-four pupils randomly assigned to one first grade were taught using the Early-to-Read i/t/a Series, and 26 pupils randomly assigned to another first grade were taught using the Ginn Basic Readers in a traditional manner. A sub-control group composed of three first grades' 86 pupils was selected randomly from the remaining first grades in the school system.

The randomization seemed effective in that no significant differences were found between the experimental and control groups as measured by the WISC and knowledge of the alphabet at the beginning of the school year, and no significant differences were found among the experimental, control and sub-control groups on the Pre-Reading Test.

Pupils in the experimental and control groups were tested in December 1964, March 1965, September 1965, and January 1966 for reading achievement using the Gray Oral Reading Tests and the vocabulary in isolation sub-test of the SRI in December 1964 and March 1965. All three groups were tested with the SAT and the SRI at the end of May 1965 and again in May 1966. The Gray Oral Reading Tests and the SRI were transliterated for administration to the experimental group in first grade and the September 1965 testing in second grade. The SAT was administered in t.o. to all groups.

Threre were no significant differences in first grade achievement as measured by the six tests of the SAT, Primary I, except in Word Recognition between the experimental and control groups. Both the experimental and control groups were superior ($p \leq 0.05$) to the sub-control group on all six tests of the SAT. The experimental group read significantly better ($p < 0.001$) than the control and sub-control groups when reading from i/t/a according to the SRI administered at the end of grade one.

There were no significant differences ($p > 0.05$) in second grade achievement between the experimental and control groups as measured by the seven tests of the SAT, Primary II, except in Arithmetic Concepts ($p < 0.01$); there were no significant differences in maximum or minimum instructional reading levels as measured by the SRI in May of second grade; there were no significant differences on the nine sub-tests of the Standard Reading Inventory except in Pronouncing Vocabulary in Isolation ($p < 0.01$). There were no significant differences ($p > 0.05$) in error pattern when reading orally on the SRI between the experimental and control groups. There were no significant differences ($p > 0.05$) in over-all reading achievement as measured by the Gray Oral Reading Test in second grade in September or January.

There were fairly consistent and significant differences ($p \leq 0.05$) between both the experimental and sub-control groups, and between the control and sub-control group on most measures of reading achievement at the end of first grade and at the end of second. The experimental group consistently achieved the highest scores and the sub-control group consistently achieved the lowest scores. There were more significant differences between the experimental and sub-control groups than there were between the control and sub-control groups.

REFERENCES

1. John A. Downing. tw bee OR NOT TO BE, Pitman Publishing Corporation, New York, 1962.
2. John A. Downing. "Experiments with Pitman's Initial Teaching Alphabet in British Schools," READING AS AN INTELLECTUAL ACTIVITY, International Reading Association Conference Proceedings, 8:191-202, Scholastic Magazines, New York, 10036, 1963.
3. Albert J. Mazurkiewicz and Harold J. Tanyzer. EARLY TO READ i/t/a PROGRAM, i/t/a Publications, Inc., New York, 1963.
4. Albert J. Mazurkiewicz. "A comparison of i/t/a and T.O. Reading, Writing, and Spelling Achievement when Methodology Is Controlled." The Initial Teaching Alphabet and the World of English, pp. 59-65, The Initial Teaching Alphabet Foundation at Hofstra University, Hempstead, N.Y., 1966.
5. Albert J. Mazurkiewicz and others. Final Newsletter, i/t/a Studies Center, Lehigh University, mimeographed, undated, received October 1966.
6. i/t/a/ BULLETIN, Initial Teaching Alphabet Publications, Inc., New York, published quarterly beginning in 1963.
7. Albert J. Mazurkiewicz. "I.T.A. and T.O. Reading Achievement when Methodology Is Controlled," The Reading Teacher, 19:606-610, May 1966.
8. Robert B. Hayes. "I.T.A. and Three Other Approaches to Reading in First Grade," The Reading Teacher, 19:627-630, May 1966.
9. Harold B. Tanyzer and Harvey Alpert. "Three Different Basal Reading Systems and First Grade Reading Achievement," The Reading Teacher, 19:636-642, May 1966.
10. Edward B. Fry. "First Grade Reading Instruction Using Diacritical Marking System, Initial Teaching Alphabet and Basal Reading System," The Reading Teacher, 19:636-642, May 1966.

11. Harry T. Hahn. "Three Approaches to Beginning Reading Instruction-I.T.A., Language Arts, and Basic Readers," The Reading Teacher, 19:590-594, May 1966.

12. Barbara Jones. "An Attempt to Extend Control Over Variables in a Second British Experiment with i/t/a." Initial Teaching Alphabet and the World of English (Albert J. Mazurkiewicz, Editor), pp. 66-78. The Initial Teaching Alphabet Foundation at Hofstra University , Hempstead, N.Y., 1966.

13. Lloyd M. Dunn, Max W. Mueller, M. D. Neely. "The Efficacy of the Initial Teaching Alphabet and The Peabody Language Development Kit with Grade One Disadvantaged Children." The Initial Teaching Alphabet and the World of English (Albert J. Mazurkiewicz, Editor) pp. 91-100, The Initial Teaching Alphabet Foundation at Hofstra University, Hempstead, N.Y., 1966.

14. William D. Sheldon, Eric F. Gardner, and Queenie B. Mills. PRE-READING TEST to accompany the SHELDON BASIC READING SERIES, form I and II. Allyn and Bacon, Inc., Boston, 1963.

15. David Wechsler. WECHSLER INTELLIGENCE SCALE FOR CHILDREN, Psychological Corporation of America, New York, 1949.

16. William S. Gray and Helen Robinson. GRAY ORAL READING TESTS, Bobbs-Merrill Company, Inc., Indianapolis, 1963. (We wish to express our appreciation for permission to transliterate these tests for use in this study.)

17. Robert A. McCracken. STANDARD READING INVENTORY, forms A and B, Pioneer Printing Co., Bellingham, Washington 98225, 1963, 1966.

18. Truman L. Kelly and others. 1964 STANFORD ACHIEVEMENT TESTS, forms W, X, Y, Z, Harcourt, Brace and World, New York, 1964.

19. J. L. Hodges, Jr. and E. L. Lehman. BASIC CONCEPTS OF PROBABILITY AND STATISTICS, pp. 303-307, Holden-Day, Inc., San Francisco, Calif., 1964.

20. Sidney Siegel. NONPARAMETRIC STATISTICS FOR BEHAVIORAL SCIENCES, pp. 111-116 and 179-184, McGraw-Hill Book Co., Inc., New York, N.Y., 1956.

21. A. L. O'Toole. ELEMENTARY PRACTICAL STATISTICS, Macmillan, (New York) 1960-1964, pp. 137-142.

22. In all the tables asterisks are used to indicate levels of probability. One asterisk (*) indicates $p \leq 0.05$, two asterisks (**) indicate $p \leq 0.01$, and three asterisks (***) indicate $p \leq 0.001$.

23. Siegel, Sidney, op.cit., pp. 96-104.

24. Siegel, Sidney, op.cit., pp. 96-104.

25. Robert A. McCracken. "A Cross-Sectional Study To Determine the Ability of Children To Transfer in Reading from Traditional Orthography to the New Augmented Roman Alphabet," The Initial Teaching Alphabet and the World of English, pp. 101-108 (A. J. Mazurkiewicz, Editor), The Initial Teaching Alphabet Foundation at Hofstra University, Hempstead, New York, 1966.

26. E. P. Bleismer and B. H. Tarborough. "A Comparison of Ten Different Beginning Reading Programs in First Grade," PHI DELTA KAPPAN, XLVI: 500-504, (June, 1965).